Trisha Espejo

CSCI335

Project 1 Phase 3

Due 5/11/21

In this project there are 6 java files

file.java – will be use to storing all the filename and content use for inverted index

HtmlProperties.java - use to save all html file name and raw content

Project2.java - main program where main method is located

SearchEngine.java - where all the work is performed

createStem.java - change the word to its root word

stemmer.java - performs the stemming

Also stemmer.java is not my code it came from the website: <https://tartarus.org/martin/PorterStemmer/java.txt>

How to compile a program?

First you have to use the commandline via Java use flags :

javac Project2.java

java Project2 -stoplist stopwords.txt -corpusdir html -isstem yes -InvertedIndex invertIndex.txt -query query.txt -pathofdir gui

The flags is not case sensitive so you do not need to worry about upper or lower case letters

You must run the java code in the following order above

first line flag for stoplist then stopwords txt file

second is the flag for corpusdir then the html folder file

third is the flag for isstem then indicate whether you would stem or not

fourth is the flag for InvertedIndex then the InvertedIndex txt file

fifth is the flag for pathofdir then gui/text/both then txt file of choice but not require unless you put text option

For the query txt:

You can put a sentence or a word of your choice that you want to query

Here the invertedindex has two option:

stemming or no stemming

to indicate stemming

you write the flag -isstem

then you write yes

to indicate no stemming

you write the flag -isstem

then you write no

if you do not indicate the stemming the default is false and will not perform stemming

example :

-isstem yes //this will enable stemming

-isstem no //this will disable stemming

The pathtodir has three option:

The first option is print the result of the query search

into a textfile of your choice to do this you do the following:

-pathtodir flag then text then textfile of your choice

-pathtodir text result.txt

the second option is it prints the result to some gui

to do this you do the following:

-pathtodir flag then gui

-pathtodir gui

the third option is it prints the result to gui and textfile of your choice

to do this you do the following:

-pathtodir flag then both then textfile of your choice

-pathtodir both result.txt

if you put both but forgot to put in the textfile it would print the result into

the result.txt

like for example you did it like this:

-pathtodir both

if you did not specify which of the three option but you put the pathtodir flag

like for example you just did this for example:

-pathtodir

it will print the result to a text file called result.txt as default

make sure that you write the flag in correct order above

order

1st stoplist

2nd directory

3rd is the stemming

4th is the inverted index

5th is the query

6th is the search result

there are no exception not following the correct order in which the flag is place would result in error

there are 4 txt needed to run the code

invertIndex.txt - outputs the inverted index

query.txt - query you want to look up

result.txt - result of query result

stopwords.txt - list of all stop words

To generate html file

copy source code of webpage and put it in folder title it html

So in running the file you must have project folder

src code is inside the project folder with all the java code in source folder

you also need to have html folder inside the project folder with all the html webpage

you also need to have invertindex.txt, query.txt, resut.txt and stopword.txt inside the project folder

There would be a total 7 txt file in the project submission

invertIndex.txt - outputs the inverted index of without stemming

query.txt - query you want to look up

result.txt - result of query result of without stemming

stopwords.txt - list of all stop words

invertIndexStem.txt - outputs the inverted index of with stemming

resultStem.txt - result of query result of with stemming